

Outcome among surviving very low birthweight infants: a meta-analysis

Gabriel J Escobar, Benjamin Littenberg, Diana B Petitti

Abstract

Because published outcome studies are the only available source of data about the morbidity among surviving very low birthweight infants (VLBW, <1500 g) a search was carried out of 1136 references in the English language. A total of 111 outcome studies were found that reported morbidity data in cohorts of VLBW infants born since 1960. The methods used and results obtained in these studies were systematically assessed. No agreement exists about the definition of study populations, descriptive statistics, or measurement of outcome. Follow up ranged from six months to 14 years. In 85 cohorts the incidence of cerebral palsy was recorded, and in 106 that of disability. Studies that followed up infants for longer time periods reported higher incidences of disability. Studies from the United States reported higher incidences of disability than those from other countries. The median incidence of cerebral palsy among all the cohorts studied was 7.7%, and the median incidence of disability was 25.0%. Despite substantial improvements in the mortality of VLBW infants, poor outcomes among survivors are common.

During the past three decades, neonatal mortality among low birthweight (LBW, birth weight <2500 g) and very low birthweight (VLBW, birth weight <1500 g) infants has decreased dramatically.¹ This has resulted in ethical and economic problems, which are most acute with respect to the outcome of surviving VLBW infants.

Reductions in neonatal mortality associated with birth weight can be confirmed by analysis of data derived from large databases made up of birth and death certificates that are geographically based and centrally organised.¹⁻⁴ In contrast, no procedure for the follow up of surviving infants exists. Morbidity in VLBW survivors can only be inferred from published reports (outcome studies) from individual nurseries or multicentre perinatal studies, but these do not always set out explicitly to evaluate overall outcome. None the less, they are the only source of information on unselected cohorts of VLBW survivors and are widely cited in publications about health policy.⁵⁻⁷

It is critical to define rigorously the outcome of VLBW survivors. Parents deserve accurate prognostic information. Individual nurseries should evaluate their performance and the associations between changes in treatments and changes in morbidity. Regionally and nationally

it is important to monitor the overall outcome of such infants so that plans for future services can be made; this is particularly important as care of VLBW infants is one of the most expensive items in paediatrics.^{5 8 9}

When assessing outcome among VLBW infants, it is useful to pose questions that parents of such survivors might pose to a paediatrician: Will my child have cerebral palsy? Will my child be disabled? Will my child have learning difficulties or need special schooling? Answers to these questions are uncertain.

In the present study we asked the following questions:

- What are the characteristics of the methods used to evaluate the database from which inferences about outcome of VLBW infants were based?
- What inferences can be made about long term morbidity among these infants?

Methods

PUBLISHED REPORTS

We attempted to locate all the outcome studies written in English that reported cohorts of unselected VLBW infants born since 1960. Only data provided in the reports were considered; no attempt was made to contact the authors for missing data. We examined computer bibliographies, reference lists in textbooks and reviews, and consulted experts.

INCLUSION CRITERIA

To be included, a study had to have been published in English in a medical journal before 1 March 1988. Unpublished studies, abstracts, reports appearing only in government publications, summaries of presentations given at medical meetings, or brief transcripts of meetings, were excluded.

The study had to have presented primary data on the outcome of a cohort or cohorts of human infants weighing ≤ 1500 g at birth and born after 31 December 1959.

The study had not deliberately excluded infants in a non-random fashion, except those with lethal congenital anomalies, those small for gestational age, or those who were products of multiple births.

The study reported the number of babies eligible for evaluation (that is, the denominator).

It had to be possible to calculate the number of unevaluated babies (those lost to follow up) in the study.

The study had to have indicated at what age outcomes were assessed.

Department of
Pediatrics,
Kaiser Permanente
Medical Center,
1425 S Main Street,
Walnut Creek,
California 94596-5300,
USA
Gabriel J Escobar

Department of Medicine,
Dartmouth-Hitchcock
Medical Center,
Hanover,
New Hampshire,
USA
Benjamin Littenberg

Division of Family and
Community Medicine,
University of California,
San Francisco,
USA
Diana B Petitti

Correspondence to:
Dr Escobar.

Accepted 1 August 1990

All infants had to have been within the specified weight range at birth.

Randomised clinical trials meeting the above criteria were eligible for inclusion.

EXCLUSION CRITERIA

The following were excluded:

Studies that did not report new data, including review articles, letters, editorials, those limiting themselves to secondary data analysis, and those examining only mortality statistics.

Studies that were limited to analysis of technical, ethical, economic, or legal aspects of the management of newborns.

Studies in which data about VLBW infants was pooled with those about infants in higher weight ranges so that the outcome of those infants weighing ≤ 1500 g could not be abstracted independently.

Studies that excluded VLBW survivors without reporting the number of infants excluded so that it was not possible to calculate the denominator.

Studies in which the time period or location of the births that were reported could not be determined.

Studies in which data about infants born before 31 December 1959 were included in such a way that the outcome of infants born after that date could not be abstracted independently.

Studies that excluded infants in a non-random fashion such that the outcome of all VLBW infants at the study site could not be determined; this group included studies that followed only a selected group (for example, only mechanically ventilated infants or only those with cerebral palsy).

Anecdotal reports, or those reporting less than 10 births.

Studies limited to outcome during the time spent in hospital after the birth or limited to follow up by laboratory or radiological testing without giving clinical information about the affected infants' functions. For example, a study that simply listed the radiological presence of intraventricular haemorrhage was excluded, whereas a study that reported such infants' functional abilities was included. Similarly, studies that listed grades of retinopathy of prematurity but that did not actually mention whether children were blind were also excluded.

We retrieved and examined 1136 references. The 161 references that met the inclusion criteria were grouped into 111 studies or 'article clusters' as some studies comprised more than one article.

ABSTRACTION OF DATA

Data were abstracted in a computer compatible form; this was accurate enough for three people abstracting the same studies to agree on 95 to 100% of all items. All data were abstracted by a trained research assistant whose work was reviewed by one of us (GJE).

REVIEW OF METHODS USED

We defined 'reporting' as providing enough

information for exact quantitation of a given description or outcome. If a study gave unquantifiable information (for example, 'many survivors . . .'), it was not considered as reporting a description or outcome. Various definitions of what constitutes cerebral palsy exist.¹⁰ We recorded how many studies reported the presence of cerebral palsy, regardless of the definition given. We also recorded whether a 'global' (complete) assessment had been made. A study could report this in any one of three ways: firstly, specific terms were used that evaluated a survivor's performance in activities of daily living (for example, studies that reported numbers of 'severely handicapped' or 'completely normal' infants), whether or not these terms were defined. The second way to present enough information for the infants to be evaluated by at least two tests; for example, some studies provided information on the Bayley test and on the presence of cerebral palsy in varying degrees of severity so that one group of infants were defined as 'normal' and another as 'abnormal'. Thirdly, some studies simply referred to another scale of handicap or functional ability (for example, 'handicapped according to Saigal's criteria'), and these were considered to have carried out a complete assessment. If a study simply reported a result on one test (for example, 32% of infants failed to perform the Goodenough draw-a-person test), we did not consider this a complete assessment. This definition of complete assessment was imposed by the heterogeneity of the studies and is in itself a major problem in reports of outcome, which make no attempt to use standard definitions of impairments, disabilities, or handicaps.¹¹ For example, not all studies consider deafness, blindness, or cerebral palsy to be handicaps, and one study's criteria of 'minor' or 'moderate' might be considered to be 'severe' or 'extreme' by another.

Studies did not provide detailed information on how long they had followed up infants, making it impossible accurately to abstract well defined cohorts followed up for discrete periods of time. Consequently we could code only the stated, minimum, and maximum follow up periods. The stated follow up period was the time given in the title, abstract, or introduction. The minimum follow up was that of the infant(s) followed up for the shortest period, and the maximum was that of the infant(s) followed up for the longest period.

Data were compared by the χ^2 test with Yates's correction for continuity or Fisher's exact test, as appropriate. All p values are reported for two tailed comparisons.

CEREBRAL PALSY AND 'DISABILITY'

A cohort was defined as a group of VLBW survivors assembled at a specified time and place and followed up from birth to determine the morbidity. Generally, each eligible report studied one or more cohorts. To be eligible for this part of the analysis, a study had to have given data about infants in such a way that the outcomes of VLBW infants followed up for at least 12 months could be assessed indepen-

dently of those followed up for less than 12 months. We abstracted data from all studies that reported incidences of cerebral palsy and from all studies that evaluated infants for the presence of a disability or handicap.

The following information was abstracted for each cohort: the weight range of the infants; whether the infants were excluded from receiving neonatal intensive care (however defined by the investigators); whether any attempt at impartial assessment took place; inclusion dates of the study; length of the minimum follow up period; definition of survivor; number of survivors; number of infants who died during the follow up period; number of infants listed as 'lost to follow up'; number of infants not evaluated; use of a geographically defined catchment area; and country of origin. In cases in which the report was ambiguous about the outcomes of some of the children studied or in which some infants were evaluated at less than 12 months of age, the infants were counted as 'not evaluated'.

In studies reporting the incidence of cerebral palsy, that for a given cohort was the number of infants with cerebral palsy divided by the total number of infants evaluated. In studies that reported complete assessments, we abstracted the number of infants that were considered 'normal', as well as how many had 'severe' or 'moderate' disabilities or handicaps, however defined. The incidence of 'disability' is thus defined as the number of infants with any disability or handicap divided by the total number of infants evaluated. We recognise that this use of the word 'disability' is not completely satisfactory.

For each subgroup of cohorts, the morbidity was determined by calculating the median and range of reported incidences, the mid-range of reported incidences (all reports between the 25th and 75th centiles that is, the middle two quartiles of the distribution of reported incidences), and the 95% confidence interval (CI) of the median.¹² The centiles and ranges show the distribution of reported results while the CI shows how close the median is to the mean. Because of a few reports of extremely high incidences, the median was usually less than the mean and therefore a more conservative estimate.

Differences between subgroups of cohorts were compared with the Mann-Whitney U test.¹³ If more than two groups were compared, the median test was used.¹⁴ The correlations between reported morbidity and continuous data (such as age at evaluation) were estimated by the Spearman rank correlation test.¹³ We chose non-parametric tests because of the skewed distribution of many of the data.

Results

THE DATA

All the 111 studies that were included were from developed nations and reported 26 099 births world wide; 12 064 of these were from the United States. The studies fell into two broad categories: 98 (88%) were multiple outcome studies that purported to provide comprehensive information on VLBW survivors. Thir-

teen (12%) were limited outcome studies that explicitly set out to examine only one specific outcome—for example, the incidence of chronic lung disease among survivors. Ninety seven studies (87%) sampled infants at a single medical centre or perinatal clinic (single centre studies). In a small subset of these (n=4, 4% of the total) two such institutions had pooled their outcome results (cooperative studies). Fourteen (13%) defined a particular catchment area and attempted to locate all the infants in that area (geographic studies); seven of these reported all infants born in a defined area, and seven reported all infants born to mothers residing in a defined area. Although geographic studies made up only 13% of the total number of studies, they accounted for a quarter of the infants. Three of the 14 geographic studies were from the United States.

DEFINITIONS AND DESCRIPTIONS OF THE STUDY POPULATIONS

Nineteen different denominators were given, including live births, admissions within 24, 48, 72, or 96 hours of birth, survivors at 36 hours, and survivors at discharge. The following upper weight limits were given: 750, 800, 999, 1000, 1250, 1360, 1499, and 1500 g. Although not all studies reported the lower weight limit in their cohorts, we found no reference to a surviving infant weighing less than 500 g. Inclusion dates for infants varied considerably; some studies reported cohorts that had been gathered over periods longer than five years. Similarly, no agreement existed on how study populations should be described (table 1).

Only 10 studies (9%) provided sufficient information—number of live births, neonatal deaths, infant deaths, and post-infant deaths—that would permit comparison with databases of vital statistics. Only 53 (48%) of the studies provided any data on the infants' socioeconomic status. In 37 of these, enough data were given for us to work out whether the cohorts were composed predominantly of infants of low or high socioeconomic status. No study provided sufficient socioeconomic information to permit adequate comparison with another study.

INFANTS TRANSPORTED FROM OTHER HOSPITALS

Twenty seven of the studies (24%) reported only babies who were born at that centre and 43

Table 1 Description of study populations in 111 eligible studies

	No (%) of studies
No of studies reporting:	
Actual No of live births	26 (23)
No of deaths in delivery room	7 (6)
No of neonatal deaths	46 (41)
No of infant deaths	49 (44)
No of deaths in hospital	37 (33)
No of late deaths	50 (45)
No of small for gestational age infants	60 (54)
Breakdown of infants' gestational ages in increments of <4 weeks' duration	15 (14)
No of infants mechanically ventilated	56 (50)
Distribution by gender	57 (51)
Distribution by race	30 (27)
No of infants transported from other hospitals	70 (63)

(39%) reported on mixed groups, the percentage born in the centre ranging from 0 to 92%. Only one study permitted comparison of morbidity between those born at the referral centre and those born elsewhere. Forty one (37%) did not report how many infants had been transported.

TRACING INFANTS AND LENGTH OF FOLLOW UP
Only 15 of the studies (14%) used databases of vital statistics or home visits to track infants. The mean minimum follow up period was 24.8 months (range 1 to 168). The mean maximum follow up period was 51.7 months (range 4 to 224). Only 23 studies (21%) followed infants up for at least 36 months, and 19 (17%) included infants follow up for less than 12 months in their reported outcome rates.

Table 2 Characteristics of cohorts of VLBW survivors

	No of cohorts	No of survivors	No of patients evaluated
Studies evaluating infants for presence of cerebral palsy (n=85)			
Birth weight range (g)			
500 (± 1) to 1500 (± 1)	45	5024	4269
999 (± 1) to 1500 (± 1)	10	1044	901
500 (± 1) to 1000 (± 1)	15	782	665
500 (± 1) to 1250 (± 1)	8	501	406
500 (± 1) to 750 (± 50)	4	119	97
Miscellaneous	3	68	61
All VLBW infants	85	7538	6399
Definition of survivor			
Not specified	11	1009	787
Survived to discharge	32	2888	2495
Survived 7 days	2	17	17
Survived 28 days	31	2609	2158
Miscellaneous (range 70–1217 days)	9	1015	942
Studies performing a complete assessment of disability (n=106)			
Birth weight range (g)			
500 (± 1) to 1500 (± 1)	49	6014	4822
999 (± 1) to 1500 (± 1)	15	1436	1260
500 (± 1) to 1000 (± 1)	28	988	830
500 (± 1) to 1250 (± 1)	8	532	449
500 (± 1) to 750 (± 50)	4	95	78
Miscellaneous	2	37	35
All VLBW infants	106	9102	7474
Definition of survivor			
Not specified	21	1256	1001
Survived to discharge	41	3463	2949
Survived 7 days	2	31	31
Survived 28 days	32	2906	2385
Miscellaneous (range 70–1217 days)	10	1446	1108

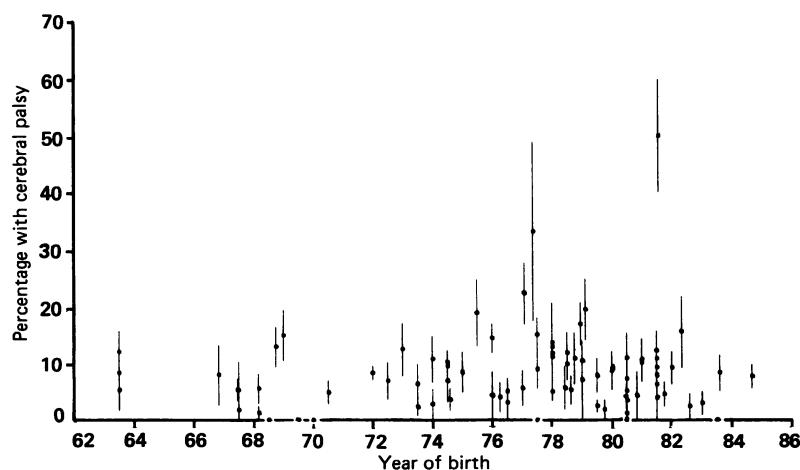


Figure 1 Incidence of cerebral palsy among VLBW survivors. The vertical axis represents the reported incidence for each of 85 eligible cohorts. The horizontal axis indicates the midpoint of the period of enrollment. Each bar represents the mean (SD) rate of a single cohort of subjects.

DETERMINATION OF OUTCOME AND REPORTED OUTCOMES

Only one study compared a parent's assessment with that of a health professional. Forty one of the studies (37%) presented data obtained from comparison groups (for example, a matched group of infants born at term). Fifteen studies (14%) compared outcome in cohorts assembled in different time periods (historical controls). Thirty two of the 98 multiple outcome studies (33%) attempted to obtain impartial assessment (masking or blinding), and two of the 13 limited outcome studies (15%) did so. These figures may overestimate the use of masking, as any attempt at masking was coded as an impartial assessment. Studies varied widely in what they considered as 'masking': in some it was the data collector who was unaware of the group that the infant was in, but the person who actually assigned infants to outcome categories was aware of the infant's clinical course.

Many important outcomes were not reported. Among the 98 multiple outcome studies, incidences of cerebral palsy were reported by 74 (76%), blindness by 70 (71%), deafness by 61 (62%), and sudden infant death syndrome by 47 (48%). Studies from the United States were less likely to report these outcomes than those from other countries (p values ranged from 0.0005 to 0.068).

We compared the 53 studies published before 1983 with the 58 published in 1983 and since. There were no significant differences between the two periods in methods of sampling or ascertainment, reporting of descriptions or outcomes, or length of follow up.

Of the 111 studies in the data set, 82 (74%) carried out complete assessments, 75 (68%) reported the incidence of cerebral palsy, and 67 (60%) did both. The 85 cohorts that reported on cerebral palsy included 7538 surviving infants of whom 6399 were actually evaluated at the age of 12 months or more; the percentages lost to follow up ranged from 0 to 53% (median 10.9%). The 106 cohorts that reported the incidence of disabilities included 9102 surviving infants of whom 7474 were actually evaluated at

Table 3 Reported incidences of cerebral palsy

Subgroup	No of cohorts	Median (%)	95% CI	Range	p Value
All cohorts	85	7.7	5.3 to 9.0	0–50.0	
*No of survivors evaluated:					
<67	44	6.6	4.4 to 10.6	0–50.0	0.48
≥67	41	8.3	5.3 to 9.4	1.1–17.1	
*No of survivors/year:					
<25	41	5.3	3.0 to 8.0	0–33.3	0.002
≥25	44	9.1	7.3 to 10.6	2.4–50.0	
Midpoint of enrollment:					
1960–1977	40	6.0	4.9 to 8.5	0–33.3	0.30
1978–1986	45	8.0	5.7 to 10.1	0–50.0	
Country of birth:					
United States	29	5.7	4.6 to 11.1	0–50.0	0.72
Other	56	7.9	5.7 to 9.0	0–33.3	
Length of follow up (months):					
<24	43	6.4	4.4 to 8.3	0–22.4	0.13
≥24	42	8.4	5.6 to 10.4	0–50.0	

*The median cohort size was 67. The median number of VLBW survivors/year was 25. Results of additional subgroup analyses available from the authors.

the age of 12 months or more; the percentages lost to follow up ranged from 0 to 55% (median 11.3%) (table 2).

CEREBRAL PALSY AND DISABILITY AMONG VLBW SURVIVORS

The reported incidences of cerebral palsy and some subgroup analyses are shown in fig 1 and table 3. The median among all the cohorts

Table 4 Reported incidences of disability

Subgroup	No of cohorts	Median (%)	95% CI	Range	p Value
All cohorts	106	25.0	20.9 to 30.0	0.0-80.0	
*No of survivors evaluated:					
<50	53	29.8	21.4 to 34.8	0.0-80.0	0.02
≥50	53	21.2	15.9 to 28.3	6.0-60.8	
*No of survivors/year:					
<23	52	30.0	21.4 to 37.4	5.9-80.0	0.02
≥23	54	21.2	17.9 to 27.3	0.0-56.8	
Midpoint of enrollment:					
1960-77	55	29.8	21.2 to 37.3	0.0-80.0	0.04
1978-86	51	21.4	18.8 to 29.3	6.0-57.7	
Country of birth:					
United States	44	30.0	25.0 to 37.9	7.1-80.0	0.002
Other	62	20.5	15.8 to 28.1	0.0-73.7	
Length of follow up (months):					
<24	57	21.5	19.6 to 27.3	0.0-57.1	0.06
≥24	49	33.6	18.8 to 39.4	5.9-80.0	

*The median cohort size was 50. The median number of VLBW survivors/year was 23. Results of additional subgroup analyses available from the authors.

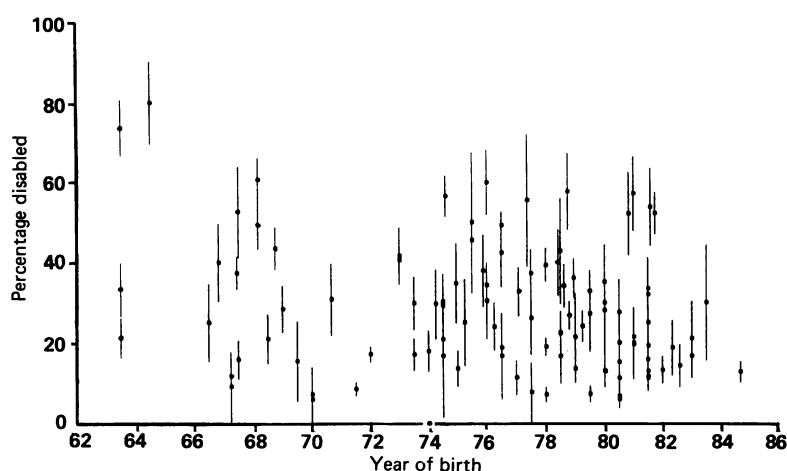


Figure 2 Incidence of disability among VLBW survivors. The vertical axis represents the reported incidence for each of 106 eligible cohorts. The horizontal axis indicates the midpoint of the period of enrollment. Each bar represents the mean (SD) rate of a single cohort of subjects.

studied was 7.7% (CI 5.3 to 9.0). There was no improvement in incidence with time (fig 1). The incidence of cerebral palsy did not differ between small and large cohorts. There were, however, significant increases in reported incidence of cerebral palsy as the number of infants who were enrolled increased (Spearman rank correlation test, $p=0.002$). Likewise, the numbers of survivors evaluated and of VLBW survivors/year (a possible marker for experience or volume of care) correlated positively with the incidence of cerebral palsy. No other differences were significant.

The median incidence of disabilities (table 4 and fig 2) among all eligible cohorts was 25.0% (CI 20.9 to 30.0). The incidences were significantly higher among the studies from the United States than those from other countries. In contrast to the results concerning cerebral palsy, incidences were lower in larger and more recent cohorts and in centres that enrolled more survivors/year. Studies that followed children up for longer periods tended to report higher disability rates. Cohorts of high socioeconomic status had lower rates of disability than those of low status but this was not significant. Cohorts for which socioeconomic data could not be extracted ($n=64$), however, had significantly lower incidences of disability than those with either high ($n=18$) or low ($n=24$) status ($p=0.011$).

Notably there was no difference between reported incidences of cerebral palsy or disabilities in cohorts of infants weighing less than 1000 g at birth and those between 1000 and 1500 g (table 5).

Tests of correlation confirmed the trends noted in the subgroup analyses of incidences of disabilities. There were significant negative correlations between the incidence of disabilities with mid-year of enrollment period, number of survivors enrolled, number of survivors evaluated, and number of survivors enrolled each year. There was a highly significant positive correlation between incidence of disability and length of follow up. Studies with the highest incidences of disability did not differ significantly in methods of reporting when compared with studies with the lowest incidences.

Discussion

In considering the reported fate of surviving VLBW infants, one must first consider how the population was defined: what was the denomi-

Table 5 Influence of birth weight on rates of cerebral palsy and disability

	Birth weight (g)			
	0-1500	1000-1500	0-1000	Other ranges
Incidence of cerebral palsy*				
No of cohorts	45	10	15	15
Median (%)	8.3	6.4	7.0	4.4
95% CI	5.7 to 10.1	1.1 to 12.9	4.4 to 10.9	2.4-9.8
Range	0-22.4	0-33.3	0-15.1	0-50.0
Incidence of disability**				
No of cohorts	49	15	28	14
Median (%)	20.9	22.6	27.4	30.0
95% CI	17.1 to 32.9	11.7 to 39.4	20.0 to 30.8	21.4-57.1
Range	0-80	6.8-60.8	9.1-60.0	14.3-73.7

Four way comparison: * $p=0.36$; ** $p=0.59$.

Table 6 Estimates of incidence of cerebral palsy in VLBW infants

Study	Birth years	Denominator	Age at follow up (years)	Percentage incidence of cerebral palsy
Stanley and Watson ²⁶	1968–81	Live births	5	1.4–2.7
Stanley and Watson ²⁶	1968–71	Survivors at 28 days	5	4.4
Stanley and Watson ²⁶	1972–74	Survivors at 28 days	5	4.6
Stanley and Watson ²⁶	1975–78	Survivors at 28 days	5	2.8
Stanley and Watson ²⁶	1979–81	Survivors at 28 days	5	3.8
Hagberg <i>et al</i> ^{28, 29}	1963–66	Survivors at 7 days	4	6.9
Hagberg <i>et al</i> ^{28, 29}	1967–70	Survivors at 7 days	4	4.5
Hagberg <i>et al</i> ^{28, 29}	1971–74	Survivors at 7 days	4	6.0
Hagberg <i>et al</i> ^{28, 29}	1975–78	Survivors at 7 days	4	5.4
Hagberg <i>et al</i> ^{28, 29}	1979–82	Survivors at 7 days	4	8.0
NCPP ²⁷	1959–66	Survivors at 1 year	7	9.0
Present study	1960–85	See text	≥1	7.7

Stanley and Watson reported a prevalence survey from Western Australia.²⁶ Hagberg *et al* reported case control studies from Sweden.^{27, 28} The National Collaborative Perinatal Project (NCPP)²⁷ was a cohort study that originally included 304 VLBW infants; results shown here are based on 188 survivors at 1 year.

nator? The incidence of any outcome is strongly dependent on the denominator used.^{10, 15} The characteristics of the population are also important, particularly with respect to the numbers of infants transferred from other hospitals^{16, 17} and the distribution of gestational ages.^{5, 18}

Once a population is defined, it is necessary to decide what outcomes are to be measured. Because the list of possible outcomes to measure is infinite, it is important to consider those conditions that will be relevant to survivors, their parents, health care providers, and planners. Though this point seems obvious, it has not always been considered in reports of outcome. At the least, one wants to know how many VLBW survivors have cerebral palsy, are blind, deaf, or will need special schooling.

The length of the follow up period is critical because some conditions present late. For example, the Western Australia cerebral palsy register has reported that 30% of cases of cerebral palsy are diagnosed after the age of 2 years.¹⁹ In the United States, Palfrey *et al* found that only 28.7% of children with special needs were identified before the age of 5 years.²⁰ These results are supported by our findings. As infants get older, the role of their socioeconomic status in determining their outcome might increase; Palfrey *et al* found that maternal education was a significant predictor of age at identification.²⁰

In addition, tracing of infants becomes more difficult over time. It is not clear what sort of infants are most likely to be lost to follow up. Tyson *et al* suggested that, at least with respect to infants in lower socioeconomic groups, untraced survivors may be equally as likely to be handicapped as those who are evaluated.²¹ Wariyar and Richmond, in a study that achieved 100% follow up, reported significantly higher disability rates among those infants who were more difficult to trace.²² Time is also important when comparing studies, because standards of care change.

Who evaluates infants is also important. Observers may be biased when evaluating infants whose medical history is known, or who they have cared for personally. Similarly, parents actually caring for handicapped infants may have different perceptions of the severity of a particular child's handicap than a neonatologist. It is to avoid such biases that most investiga-

tors consider it important to ensure impartial observers.^{10, 23, 24}

Finally, once the incidence of a given outcome is determined, it is important to ask what this means and how it relates to the general population. It is important to compare outcomes with those of term infants,²³ as there are published data on the prevalence of such outcomes in the general population.^{10, 19, 25–31}

The following conclusions can be made about published reports about the outcome for VLBW infants:

(i) The database is small. We could only locate reports on 26 000 babies in almost 30 years—but each year almost 40 000 are born in the United States alone. The number of infants actually evaluated for outcomes such as cerebral palsy was even smaller, roughly 7000 infants. Even fewer infants were followed up until school age, and the number of infants born weighing less than 1000 g who were evaluated was less than 1000.

(ii) There was no agreement on what should be the appropriate denominator. This not only limited the value of comparisons between studies, but also precluded comparison with databases of vital statistics, which are based on numbers of live births.

(iii) Although morbidity among survivors is often at the centre of ethical debates, it is striking that there is no agreement as to what constitutes an adverse outcome. Conditions such as blindness, deafness, and cerebral palsy were not always reported or were considered to be of minor importance.

(iv) Most of the published reports consist of data on infants followed up for less than three years. Furthermore, because of inconsistent reporting, little attention paid to preventing selection or ascertainment biases, and little use of impartial observers, it seems likely that the reports underestimate the morbidity among these infants.

(v) Little improvement in the methods of carrying out these studies has taken place over time.

With respect to the actual outcome of VLBW survivors, the following inferences can be made:

(i) The most striking aspect of the papers that we reviewed was the tremendous variation among the outcomes reported, with studies in the same time period reporting widely divergent

results. This heterogeneity was similar to that noted among internursery comparisons of mortality and incidences of chronic lung disease.³²⁻³⁴

(ii) The median incidence of cerebral palsy reported among all the cohorts studied was 7.7% and showed little variation over time; this figure is similar to those given in other studies (table 6). If one assumes the incidence of cerebral palsy in the general population to be 2/1000 live births or neonatal survivors (a generally accepted estimate²⁵) then the relative risk for cerebral palsy among surviving VLBW infants would be 38 times that in the general population.

(iii) Our best estimate for the incidence of disability in VLBW survivors in these reports is about 25%. Because studies did not uniformly report adverse outcomes, particularly lung disease, this figure may be an underestimate. Furthermore, we think that evaluation at too young an age may be resulting in an incorrectly low incidence being reported. As other birthweight specific databases do not exist, external confirmation or refutation of this estimate is currently impossible. The reported incidence of disability varies from centre to centre, and is related to the size of the study, the age at which the children are evaluated, and the date of the midpoint of a study's enrollment period.

(iv) Also striking is the increased incidence of disability reported in studies that followed up infants for longer periods of time, an issue raised by other investigators.³⁵ This finding is all the more remarkable when one considers the crudity of our measure for length of follow up: for this analysis, we could code only the minimum follow up period for each cohort.

(v) The surprising finding that studies from the United States report higher incidences of disabilities, though statistically robust, awaits independent validation in studies using uniform criteria for the denominators and outcome measures.

THE LIMITATIONS OF THIS ANALYSIS

Many of the subgroup analyses were designed to assess impact of study design on reported outcome. It is surprising to us that morbidity did not differ with the use of geographic inclusion criteria or of impartial evaluators, or that it did not differ among cohorts of different birth weights. Many of our subgroup analyses had only borderline significance. Furthermore, the fact that we tested multiple comparisons means that the calculated *p* values may be misleading. As we performed 16 tests of significance on each outcome, a conservative analysis would set the threshold for significance at $0.05/16=0.003$. In this case, only our strongest findings among subgroups (that studies from the United States reported higher incidences of disability than those from other countries and that centres with larger numbers of VLBW infants report higher incidences of cerebral palsy) would be called 'statistically significant'.

The most important limitation, however, is in the data themselves: the figures that we have derived can only be considered crude estimates

because of the limitations inherent in the studies themselves. It is important to remember that, even to make these crude estimates, it was necessary to make a number of simplifications—such as assuming that one can pool results based on 28 day survivors with those using survivors of unspecified duration. Furthermore, the imprecision of the methods used in the studies undoubtedly affected our results. For example, when considering variations in reported incidences of disability over time one must keep in mind that the studies reported enrollment periods ranging from 8 months to 16 years.

The striking finding of a 50% greater rate of disability among babies studied in the United States compared with those born in other countries does not seem to be the result of differences in the definitions of disability. We cannot discern from our data if this association is the result of differences in health care technology or delivery around the world, differences in the definition of disability across countries, or differences in the biology or mortality of VLBW in the different centres. Nurseries in the United States may be more aggressive than those in other countries in treating the highest risk infants, and thus be prone to greater morbidity.³⁶

Before ascribing too much meaning to this finding, however, one must consider some limitations in our analysis. Firstly, we excluded studies not published in English. Secondly, our estimates were based on a small number of infants and spanned several decades. Further refinement of our estimates using available data would be difficult. For example, one might wish to include only studies that included babies born after 1980—a reasonable approach in the light of recent changes in the care of these infants. Only 21 cohorts would be available, with a total of 1408 babies evaluated. Should one generalise to the population of VLBW infants at large after studying only 1408 infants?

Moreover, whereas some comparison is possible with other studies of the incidence of cerebral palsy, no such comparison is possible of handicap, as other birthweight specific databases do not exist. For example, a frequently cited source on morbidity among all children in the United States (the Child Health Supplement of the National Health Interview Survey) includes only 90 VLBW infants, all of whose weights were determined by maternal recall.³⁷ A recent meta-analysis of LBW outcome studies focused on intelligence and developmental quotients, and did not mention the question of what the incidence of cerebral palsy or disability was among these infants.³⁸

Recommendations

(1) Evaluation of the outcome of VLBW survivors should be given priority in paediatric research.

(2) National paediatric and medical associations, as well as the editorial boards of important medical journals, should establish minimum standards for the design, conduct, and reporting of outcome of VLBW survivors. Standards should be set that deal with the

concerns about different methods that have been raised in this analysis and by other investigators.^{6 17 21 38 39}

(3) National paediatric and medical associations should encourage the development of geographically based birth weight specific morbidity databases so monitoring of the overall outcome of VLBW infants is possible.

(4) Outcome studies should be prospective, use impartial observers, include full term infants as controls, and include assessments of the parents of VLBW survivors.

(5) Agreement on defining outcome measures, even crude ones, must be reached so that studies can be compared. The experience of population based cerebral palsy registries should be used as models.

(6) Institutional incentives and funding must be offered so that follow up of VLBW survivors, particularly multicentre and geographical studies, is financially and professionally feasible. Both the confidentiality of patients and the protection of physicians from allegations of malpractice must be assured by each institution.

Despite the increased survival of VLBW infants, it is extremely difficult to know how the survivors are faring. This problem has been ignored in discussions about whether individual infants should receive 'heroic' care or whether the '800 gram barrier' can be broken.⁵ Our ability to assess and report the outcome of surviving VLBW infants has lagged behind our willingness to resuscitate them.

The authors are indebted to Drs Ronald Ariagno, Iain Chalmers, Lucy Crain, Alan M Garber, Halsted R Holman, Lincoln E Moses, Nigel Paneth, Marie C McCormick, Saroj Saigal, William A Silverman, John C Sinclair, Marc Usatin, and Ted Zukin for their helpful suggestions. Gloria Linder of Stanford University's Lane Medical Library provided invaluable assistance in our search of published reports. We particularly thank Lynn S Gordon for her help in abstracting the data.

Appendices giving details of the search of publications, all references, and expanded tables with analyses of subgroups are available from Dr Escobar. Dr Escobar and Dr Littenberg were fellows at the Stanford University Medical Center Clinical Scholars Program, and the work was supported by the Robert Wood Johnson Foundation and the Kaiser Family Foundation.

- 1 McCormick MC. The contribution of low birth weight to infant mortality and childhood morbidity. *N Engl J Med* 1985;312:82-90.
- 2 Wegman ME. Annual summary of vital statistics: 1988. *Pediatrics* 1989;84:943-56.
- 3 Williams RL. Measuring the effectiveness of perinatal medical care. *Med Care* 1979;17:95-109.
- 4 Binkin NJ, Rust KR, Williams RL. Racial differences in neonatal mortality: what causes of death explain the gap? *Am J Dis Child* 1988;142:434-40.
- 5 Office of Technology Assessment. *Neonatal intensive care for low birthweight infants: costs and effectiveness*. Washington, DC: US Government Printing Office. Health Technology Case Study 38, December 1987.
- 6 Ozminkowski RJ, Wortman PM, Roloff DW. Evaluating the effectiveness of neonatal intensive care. What can the literature tell us? *Am J Perinatol* 1987;4:339-47.
- 7 Budetti PP, McManus P. Assessing the effectiveness of neonatal intensive care. *Med Care* 1982;20:1027-39.
- 8 Hernandez JA, Offutt J, Butterfield LJ. The cost of care of the less-than-1000-gram infant. *Clin Perinatol* 1986;13:461-76.
- 9 Shankaran S, Cohen SN, Linver M, Zonia S. Medical care costs of high-risk infants after neonatal intensive care: a controlled study. *Pediatrics* 1988;81:372-8.
- 10 Stanley FJ, Alberman E. *The epidemiology of the cerebral palsies*. Philadelphia: Spastics International Medical Publications, 1984.
- 11 World Health Organisation. *International classification of impairments, disabilities, and handicaps*. Geneva: World Health Organisation, 1976.
- 12 Snedecor GW, Cochran WG. *Statistical methods*. 7th Ed. Ames: Iowa State University Press, 1980.
- 13 Siegel S. *Nonparametric statistics for the behavioural sciences*. New York: McGraw-Hill, 1956.
- 14 Brown GW, Mood AM. On median tests for linear hypotheses. In: Neyman J, ed. *Proceedings of the second Berkeley symposium on mathematical statistics and probability*. Berkeley, University of California Press, 1951:159-66.
- 15 Kitchen WH, Murton LJ. Survival rates of infants with birth weight between 501 and 1000 g. Improvement by excluding certain categories of cases. *Am J Dis Child* 1985;139:470-1.
- 16 Paneth N, Kiely JL, Susser M. Age at death used to assess the effect of interhospital transfer of newborns. *Pediatrics* 1984;73:854-61.
- 17 Ozminkowski RJ, Wortman PM, Roloff DW. Inborn/outborn status and neonatal survival: a meta-analysis of non-randomized studies. *Stat Med* 1988;7:1207-21.
- 18 Verloove-Vanhoorick SP, Verwey RA, Brand R, Gravenhorst JB, Keirse MJNC, Ruys JH. Neonatal mortality risk in relation to gestational age and birthweight. *Lancet* 1986;ii:55-7.
- 19 Stanley FJ, Watson L, Mauger S. *Second report of the western Australia cerebral palsy register*. Nedlands, Australia: National Health and Medical Research Council Research Unit in Epidemiology and Preventive Medicine, 1987.
- 20 Palfrey JS, Singer JD, Walker DK, Butler JA. Early identification of children's special needs: a study in five metropolitan communities. *J Pediatr* 1987;111:651-9.
- 21 Tyson JE, Laskey RE, Rosenfeld CR, Dowling S, Gant N. An analysis of potential biases in the loss of indigent infants to follow-up. *Early Hum Dev* 1988;16:13-25.
- 22 Wariyar UK, Richmond S. Morbidity and preterm delivery: importance of 100% follow-up. *Lancet* 1989;ii:387-8.
- 23 Kiely JL, Paneth N. Follow-up studies of low-birthweight infants: suggestions for design, analysis and reporting. *Dev Med Child Neurol* 1981;23:96-100.
- 24 Fletcher RH, Fletcher SW, Wagner EH. *Clinical epidemiology: the essentials*. Baltimore: Williams and Wilkins, 1988.
- 25 Paneth N, Kiely JK. The frequency of cerebral palsy: a review of population studies in industrialized nations since 1950. In: Stanley FJ, Alberman E, eds. *The epidemiology of the cerebral palsies*. Philadelphia: Spastics International Medical Publications, 1984.
- 26 Stanley FJ, Watson L. The cerebral palsies in western Australia: trends 1968 to 1981. *Am J Obstet Gynecol* 1988;158:89-93.
- 27 Ellenberg JH, Nelson KB. Birth weight and gestational age in children with cerebral palsy or seizure disorders. *Am J Dis Child* 1979;133:1044-8.
- 28 Hagberg B, Hagberg A, Olow I. The changing panorama of cerebral palsy in Sweden. IV. Epidemiological trends 1959-78. *Acta Paediatr Scand* 1984;73:433-40.
- 29 Hagberg B, Hagberg A, Olow I, Von Wendt L. The changing panorama of cerebral palsy in Sweden. V. The birth year period 1979-82. *Acta Paediatr Scand* 1989;78:283-90.
- 30 Newacheck PW, Budetti PP, Halfon N. Trends in activity-limiting chronic conditions among children. *Am J Public Health* 1986;76:178-84.
- 31 Newacheck P, Halfon N, Budetti PP. Prevalence of activity-limiting chronic conditions among children based on household interviews. *J Chronic Dis* 1986;39:63-71.
- 32 Avery ME, Tooley WH, Keller JB, et al. Is chronic lung disease in low birth weight infants preventable? A survey of eight centers. *Pediatrics* 1987;79:26-30.
- 33 Kraybill EN, Bose CL, D'Ercole J. Chronic lung disease in infants with very low birth weight. *Am J Dis Child* 1987;141:784-8.
- 34 Horbar JD, McAuliffe TL, Adler SM, et al. Variability in 28 day outcomes for very low birth weight infants: an analysis of 11 neonatal intensive care units. *Pediatrics* 1988;82:554-9.
- 35 McCormick MC. Long-term follow-up of infants discharged from neonatal intensive care units. *JAMA* 1989;261:1767-72.
- 36 Guillemin JH, Holmstrom LL. *Mixed blessings. Intensive care for newborns*. New York: Oxford University Press, 1986.
- 37 US Department of Health and Human Services. *Child health supplement, 1981 health interview survey*. Publication No PHS 82-1569. Hyattsville: National Center for Health Statistics, 1982.
- 38 Aylward GP, Pfeiffer SI, Wright A, Verhulst SJ. Outcome studies of low birth weight infants published in the last decade: a meta-analysis. *J Pediatr* 1989;115:515-20.
- 39 Mutch LMM, Johnson MA, Morley R. Follow up studies: design, organisation, and analysis. *Arch Dis Child* 1989;64:1394-402.